Austin Publishing Group

Review Article

Finite Mixture Models and Their Applications: A Review

Hanze Zhang and Yangxin Huang*

Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, USA

*Corresponding author: Yangxin Huang, Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, Tampa, Florida 33612, USA

Abstract

Finite Mixture (FM) models have received increasing attention in recent years and have proven to be useful in modeling heterogeneous data with a finite number of unobserved sub-population. It has been not only widely applied to classification, clustering, and pattern identification problems for independent data, but could also be used for longitudinal data to describe differences in trajectory among these subgroups. However, due to the computational convenience, the most types of FM models are based on the normality assumption which may be violated in certain real situations. Recently, FM models with non-normal distributions, such as skew normal and skew t-distribution, have received increasing attention and showed the advantages in modeling data with non-normality and heavy tails. One of the advantages of FM models is that both maximum likelihood method and Bayesian approach can be applied to not only estimate model parameters, but also evaluate probabilities of subgroup membership simultaneously. We present a brief review of FM models for these two types of data with different scenarios.

Keywords: Finite mixture models; Heterogeneity; Longitudinal data; Non-normal distributions

Introduction

Over 100 years ago, the famous biometrician Pearson [1] helped his colleague solve a problem in accommodating apparent skewness of crab sample adequately by one symmetric normal distribution. With a strong feeling of that this population was evolving toward two new subspecies, he fitted a mixture of two normal probability density functions with different means and variances in two proportions. After Pearson firstly proposed the word "mixture" in statistics, not surprisingly, various attempts were conducted to dig deeper in this field.

Most of the statistical models assume that a sample of observations comes from the same distribution. Sometimes, however, it may not be true, since the sample may be drawn from numbers of distinct populations in which the populations are not identified. In this situation of homogeneity assumption violated, Finite Mixture (FM) models could bring the rescue. FM models provide a flexible frame work to handle heterogeneous data with a finite number of unobserved sub-population, and also have been widely applied to classification, clustering, and pattern identification problems [2-5]. FM models have attracted considerable research interest recently and have been widely applied to independent data. Recently, the use of FM models for longitudinal data has also received increasing attention. This article is organized to provide a brief overview of FM models for these two types of data with different scenarios.

Models with normal mixture for independent data

When a FM model is a convex combination of two or more probability density functions, it can be formally written as a mixture with K component distributions:

$$f(x) = \sum_{k=1}^{K} w_k f_k(x) \tag{1}$$

Where $w_k > 0$ (k=1, 2, …, K) is the mixing weights with $\sum_k w_k = 1$.

In modeling independent data, FM models allow for parameter differences across the unobserved classes. In other words, $f_k(x)$ in (1) are all from the same parametric family, but with different parameters. Many distributions have been applied as the parametric family of the components in the mixture model.

Due to the computational convenience, normal distributed components have been widely used [6]. It could be easily fitted iteratively by Maximum Likelihood (ML) via the Expectation-Maximization (EM) algorithm [6-8]. Briefly, EM algorithm includes the following steps: 1) start with initial values about the mixture components and the mixing weights $w_1, … w_k$; 2) use the current parameter guess, calculate the weights (E-step), then use the current weights, maximize the weighted likelihood to obtain new parameter estimates (M-step); 3) repeat steps 1) and 2) iteratively until convergence of algorithm, and then return the final parameter estimates and component probabilities. Several researchers have published program for the parameter estimation of FM models using EM algorithm [9-12]. Additionally, FM model has also been studied from a semi parametric prospective [13,14]. In terms of its flexibility, FM model with normal distribution has been widely applied in different areas, including, but not limited to, medicine [15-17], genetics [18-21], public health [22], psychology [23,24], and economics [25,26].

Models with normal mixture for longitudinal data

Although most of the FM models focus mainly on independent data, mixtures have also been developed for modeling longitudinal data, where the latent classes corresponding to the components and individual clusters provide a better data fitting. It aimed at identifying multiple unobserved sub-groups, and describing differences in longitudinal change among these subgroups.

Generally, the density function of FM models for longitudinal

data can be written as

$$f(\boldsymbol{y}_i) = \sum_{k=1}^{K} w_k f_k (g_k(\beta_{ik}, x_{ik}); \ \sigma_k^2(\beta_{ik}, x_{ik})) \qquad (2)$$

where $\boldsymbol{y}_i$ denotes a vector of repeated observations for subject i which is assumed to be from the kth component; $f_k(g_k(\beta_{ik}, x_{ik}); \ \sigma_k^2(\beta_{ik}, x_{ik}))$ is the density function for k^{th} component with a mean function $g_k(.)$ and variance function $\sigma_k^2(.)$; β_{ik} and x_{ik} denote unknown subject specific parameters and known covariates, respectively. Similarly, $w_k > 0$ are the mixing weights with $\sum_k w_k = 1$. For $g_k(.)$, both linear (polynomial) and non-linear mean function could be applied, but former one is more widely used, partially because the inference process can be conveniently carried out by ML approach [27]. While formularizing the FM models for longitudinal data, the mean functions of components can be similar forms with varying means and/or variance specifications, or have totally different mean trajectories across the components [28].

FM models for longitudinal data, also named as growth mixture models, were presented by Verbeke [29] and Muthen [27]. Growth mixture model is built up by combining the random effect from mixed effects models and finite mixtures, which allows same mean function but with different sets of parameter values (growth factors) across components capturing latent trajectory classes with different curve shapes [30-32]. It could be considered as an extension of the conventional Linear Mixed-Effect (LME) model with different latent classes of development. Both EM algorithm [27] and Bayesian methods [33,34] were used for estimating both model parameters and subclass membership probabilities. The relative developments, called latent class growth analysis [35-37], were special cases that assume no inter-individual differences in change within-class. In other words, it specifies that all individuals in one trajectory class behave the same, which allows more straightforward interpretations.

All of the mixtures above had the assumption of normally distributed variables within each latent class. According to computational convenience of the normality assumption, many extensions and applications have been presented in different fields, such as medicine [38-40], psychology [41,42], social science [43-45] and pharmacokinetic/pharmacodynamic [46].

Models with non-normal mixture for independent data

In many real situations, however, the data contain longer than normal tails or atypical observations, the use of normal components may affect the fit of the model and, in turn, lead to biased results. The FM model of t-distribution was considered as an alternative, which provides a more robust approach of fitting mixtures and computes less extreme estimates of the posterior probabilities of the component membership [47-50]. It has proven to accommodate outliers in modeling data with heavy tails by an additional parameter, the degrees of freedom, compared to that with normal distribution. Expectation-Conditional Maximization (ECM) algorithm [47,48] and Bayesian approach [51] were used to fit the FM models with t-distribution. In practice, FM model with t-distribution has been implemented to wide fields, including genetics [52,53], medicine [54-56] and engineering [57].

In addition to feature of heavy tails, in many applied problems, data commonly involve highly asymmetric feature. The FM models with symmetric distributions, such as normal and t-distributions

can be misleading when handling data with skewness. Recently, asymmetric distribution-based mixture models, particularly, the Skew-Normal (SN) [58-63] and Skew-t (ST) mixture models [62,64-67] have received increasing attention and been developed as a critical extension to traditional models with symmetric distributions for modeling data with asymmetry, heavy tails, and the presence of outliers.

The FM models of SN distribution can provide a more appropriate density estimation to fit the asymmetric observations by adding an additional shape/skewness parameter, compared to the normal mixtures. Model fitting could be conducted by both EM algorithm [58,59] and Bayesian approach using Markov Chain Monte Carlo (MCMC) method [58,62]. Its flexibility and robustness against skewness has been proven in the real data, such as genetic data [68], transportation data [69], and environmental data [70].

As a natural extension of the student t and skew normal mixtures, FM model with ST distribution has showed its advantages in modeling data with both asymmetry and heavy tails simultaneously. Compared to SN and student t distribution, the ST distribution has extra parameters, degrees of freedom and shape/skewness parameter. Therefore, FM models with normal, student-t and SN distributions can be statistically viewed as special cases of the ST mixture models. Lee and McLachlan [71-73] suggested that the existing ST distributions could be classified into four forms, including restricted, unrestricted, extended and generalized forms. The EM algorithm was used for fitting mixtures of both restricted and unrestricted ST distribution [65,71,74]. The unrestricted ST mixture model has a more general characterization than various mixture models of restricted ST mixture model, and hence is able to regulate the asymmetric behaviors across components with greater flexibility [71]. A Bayesian approach implemented by MCMC scheme could also be applied to make inference for FM models with ST distribution in great efficiency [62]. Its application was found in various areas, including biology [66,75], bioinformatics [76], transportation [69] and astrophysics [77].

Other than these distributions above widely used in FM models, some alternative non-normal distributions have also received some attention, including normal inverse Gaussian distribution [78,79], skew t-normal distribution [80], Shifted Asymmetric Laplace (SAL) distribution [81], and generalized hyperbolic distributions [82]. Franczak [81] suggested that the SAL mixture models offered near-perfect results on the data whereas the mixture models with normal distribution consistently overestimated the number of components.

Models with non-normal mixture for longitudinal data

Similar to the FM model for independent data discussed above, when the repeated observations, $\boldsymbol{y}_i$ in (2), are truly non-normally distributed, the model with normal assumption is not robust and can lead to poor estimation and inference [83]. In this case, non-normal FM models for longitudinal data should be considered, because it fits the data better than normal mixture. Although most of non-normal distributions such as SN [84] and ST distributions [85] used on FM model for independent data could be applied on longitudinal cases, ST distribution was most widely implemented by adding a skew parameter and degrees of freedom parameter. Either the random effects or the residual of the model could be assumed an ST distribution. Recently, for example, Muthén [85] introduced a new growth mixture model with ST distributed random effects.

In addition to FM models with linear (polynomial) or piecewise linear mean functions, the mixture models with different nonlinear mean components have obtained increasingly attention. For instance, to explicitly estimate the HIV viral load trajectories, Huang et al. [86] constructed three different mean functions for three potential subgroups with ST distribution, including one-compartment model with a constant decay rate, two compartment model with constant decay rates, and two compartment model with constant and time-varying decay rates, respectively, and made inference for the ST-FM models from Bayesian prospective. Furthermore, in addition to non-normality, Huang et al. extended FM models by considering other longitudinal data features simultaneously, including measurement errors in covariates [87-89], non-ignorable missing mechanism [87,89-91], left-censored response [92], and time-to-event outcomes [93].

Discussion

Recent decades, FM model has proved to be one of the most powerful model-based approaches dealing with data in the presence of population heterogeneity. This heterogeneity could be detected by visual methods, such as scatter plot and histogram. For instance, a bimodal or even multi-modal distribution for independent data and distinct trajectories for longitudinal data strongly suggest the existence of heterogeneity or subgroups. FM models could handle this data feature not only by providing model parameter estimates, but also allowing estimate of model-based probabilistic clustering to obtain class membership probabilities. Recent developments and extensions in FM models offer increasing ability and flexibility in capturing independent or longitudinal data with different data features, which can benefit applications in various scientific areas.

The optimal number of mixture components selection is an important but difficult problem in FM models. Since the conventional likelihood-ratio test comparing k and k+1 components FM models is not appropriate, adjusted Lo-Mendell-Rubin Likelihood-Ratio Test (Adjusted LRT) obtained the agreement in selecting the model with optimal number of components [94]. An alternative approach to determine the optimal number of components is to compare the information criteria, such as Akaike's Information Criteria (AIC) [95], Bayesian Information Criteria (BIC) [96], and Sample-Size Adjusted BIC (SSABIC) [97]. However, most of these criteria are very sensitive to sample size, and favor highly parameterized models. Thus, it is suggested that these information criteria should be considered with other evidence. Additionally, entropy has also been considered as a criterion for components number selection. Entropy assesses weather one subject was classified neatly into one and only one subgroup, with higher value (> 0.80) indicating better classification [98]. As this issue has not been completely resolved, it is good to apply different criteria simultaneously to determine the optimal number of components for FM models.

As a constrained exploratory technique, FM model seeks the patterns that data are trying to tell, but what can be learned is limited by what is entered. In other words, the final model is the best representation of the data, given the specifications of the model before the estimation algorithm. Whether they represent the true heterogeneous patterns is unknown. Thus, we suggest researchers to obtain further evidence that the unobserved subgroups really exist by replicating findings with another data, and identifying the association between subgroup membership and other measured variables.

Initial value selection and convergence issue often appear in model estimation via EM or ECM algorithm for computationally intensive FM models. With general form of skew distributions, sometimes it may not be able to get closed form for the conditional expectations involved in the E-step of the EM algorithm. Starting with different sets of initial values is strongly recommended, which helps determine whether these values all result in the same solution. Non-normal distributed mixtures need more random initial values than normal mixtures to replicate the best log-likelihood given a typically less smooth likelihood function. To avoid these problems happened in EM or ECM algorithms, Bayesian approach with MCMC technique, which has attracted the attention in this field, could be a rescue.

Other cautions of FM models should also be addressed. First, the computational load of complicated FM models, especially mixtures with non-normal distributions for longitudinal data, is extremely heavy. Second, for inference of FM models, parameter (or model) identifiability can be a critical but difficult problem when a large number of model parameters must be estimated simultaneously. Each component model must be ensured to be identifiable, and then the whole mixture model could be identifiable. If the model is not properly identified, it is possible that many different sets of parameter estimates would appear. Moreover, models comparison and goodness fit tests need to be further developed, not only focusing on the difference in the number of latent classes, but also in their random-effects specification. Finally, FM model is a statistical procedure which is usually based on large sample size.

In summary, FM model is a fast developing statistical approach for modeling independent or longitudinal data with heterogeneity. This article provides an up-to-date brief overview of the developments in FM models for both independent and longitudinal data. Compared to independent data, studies on FM model for complicated longitudinal data are still relatively limited, and few studies include time-varying predictors, but we believe that more and more important and interesting results in this area will be reported in the near future.

A final note that we would like to make is possible software to implement FM models. The most widely used software for FM models are EMMIX [99] and Mplus [100]. Other available software designed for certain specific situations include, but not limited to, AUTOCLASS [101], NORMIX [102], and MIX [103]. Several R packages are also available to implement FM mixture models, including 'mclust' [104], 'mixtools' [105], 'FlexMix' [106]. When the mean functions of components are very complicated, especially for longitudinal data with non-normal distributions, which bring extremely heavy computational load, the Bayesian method shows its advantages. The WinBUGS software [107] interacted with the package 'R2WinBUGS' in R is a good choice.

References

1. Pearson K. Contributions to the mathematical theory of evolution. Philosophical Transactions of the Royal Society of London A. 1894: 71-110.

2. Everitt BS, Hand DJ. Finite mixture distributions: Springer. 1981.

3. Titterington D, Smith A, Makov U. Statistical analysis of finite mixture distributions. John Wiley & Sons Ltd, Chichester. 1985.

4. McLachlan GJ, Basford KE. Mixture models. Inference and applications to clustering. Statistics: Textbooks and Monographs, New York: Dekker. 1987; 1.

5. Lindsay BG. Editor Mixture models: theory, geometry and applications. NSF-CBMS regional conference series in probability and statistics. 1995: JSTOR.

6. McLachlan G, Peel D. Finite mixture models: John Wiley & Sons. 2004.

7. McLachlan G, Krishnan T. The EM Algorithm and Extensions. Wiley, New York. 1997.

8. Nityasuddhi D, Böhning D. Asymptotic properties of the EM algorithm estimate for normal mixture models with component specific variances. Computational statistics & data analysis. 2003; 41: 591-601.

9. Agha M, Ibrahim M. Algorithm AS 203: maximum likelihood estimation of mixtures of distributions. Applied Statistics. 1984; 33: 327-332.

10. Agha M, Branker D. Algorithm AS 317: Maximum Likelihood Estimation and Goodness-of-fit Tests for Mixtures of Distributions. Journal of the Royal Statistical Society: Series C (Applied Statistics). 1997; 46: 399-407.

11. Böhning D, Dietz E, Schlattmann P. Recent developments in computer-assisted analysis of mixtures. Biometrics. 1998; 54: 525-536.

12. Jones P, McLachlan G. Algorithm AS 254: Maximum Likelihood Estimation from grouped and truncated data with finite normal mixture models. Applied statistics. 1990; 39: 273-282.

13. Hunter DR, Wang S, Hettmansperger TP. Inference for mixtures of symmetric distributions. The Annals of Statistics. 2007; 35: 224-251.

14. Bordes L, Mottelet S, Vandekerkhove P. Semiparametric estimation of a two-component mixture model. The Annals of Statistics. 2006; 34: 1204-1232.

15. Schlattmann P. Medical applications of finite mixture models: Springer. 2009.

16. Schlattmann P, Bohning D. Mixture models and disease mapping. Stat Med. 1993; 12: 1943-1950.

17. Conway KS, Deb P. Is prenatal care really ineffective? Or, is the 'devil' in the distribution? J Health Econ. 2005; 24: 489-513.

18. Pan W, Lin J, Le CT. A mixture model approach to detecting differentially expressed genes with microarray data. Funct Integr Genomics. 2003; 3: 117-124.

19. McLachlan GJ, Bean RW, Jones LB. A simple implementation of a normal mixture approach to differential gene expression in multiclass microarrays. Bioinformatics. 2006; 22: 1608-1615.

20. Zhang S. An improved nonparametric approach for detecting differentially expressed genes with replicated microarray data. Stat Appl Genet Mol Biol. 2006; 5.

21. Mori K, Oura T, Noma H, Matsui S. Cancer outlier analysis based on mixture modeling of gene expression data. Comput Math Methods Med. 2013; 2013: 693901.

22. Fahey MT, Ferrari P, Slimani N, Vermunt JK, White IR, Hoffmann K, et al. Identifying dietary patterns using a normal mixture model: application to the EPIC study. J Epidemiol Community Health. 2012; 66: 89-94.

23. Mun EY, von Eye A, Bates ME, Vaschillo EG. Finding groups using model-based cluster analysis: Heterogeneous emotional self-regulatory processes and heavy alcohol use risk. Developmental Psychology. 2008; 44: 481-495.

24. Steinley D, Brusco MJ. Evaluating mixture modeling for clustering: recommendations and cautions. Psychol Methods. 2011; 16: 63-79.

25. Deb P, Trivedi PK. Demand for medical care by the elderly: a finite mixture approach. Journal of applied econometrics. 1997; 12: 313-336.

26. Deb P, Gallo WT, Ayyagari P, Fletcher JM, Sindelar JL. The effect of job loss on overweight and drinking. J Health Econ. 2011; 30: 317-327.

27. Muthén B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics. 1999; 55: 463-469.

28. Pauler DK, Laird NM. A mixture model for longitudinal data with application to assessment of noncompliance. Biometrics. 2000; 56: 464-472.

29. Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in the random-effects population. Journal of the American Statistical Association. 1996; 91: 217-221.

30. Muthén B. Latent variable analysis. The Sage handbook of quantitative methodology for the social sciences Thousand Oaks, CA: Sage Publications. 2004: 345-368.

31. Li F, Duncan TE, Duncan SC, Acock A. Latent growth modeling of longitudinal data: A finite growth mixture modeling approach. Structural Equation Modeling. 2001; 8: 493-530.

32. Wang M, Bodner TE. Growth Mixture Modeling Identifying and Predicting Unobserved Subpopulations With Longitudinal Data. Organizational Research Methods. 2007; 10: 635-656.

33. Elliott MR, Gallo JJ, Ten Have TR, Bogner HR, Katz IR. Using a Bayesian latent growth curve model to identify trajectories of positive affect and negative events following myocardial infarction. Biostatistics. 2005; 6: 119-143.

34. Asparouhov T, Muthén B, editors. Using Bayesian priors for more flexible latent class analysis. Proceedings of the 2011 Joint Statistical Meeting, Miami Beach, FL. 2011.

35. Nagin DS. Analyzing developmental trajectories: a semiparametric, group-based approach. Psychological methods. 1999; 4: 139-157.

36. Nagin D. Group-based modeling of development: Harvard University Press. 2009.

37. Sterba SK, Bauer DJ. Predictions of individual change recovered with latent class or random coefficient growth models. Structural Equation Modeling: A Multidisciplinary Journal. 2014; 21: 342-360.

38. Lin H, McCulloch CE, Turnbull BW, Slate EH, Clark LC. A latent class mixed model for analyzing biomarker trajectories with irregularly scheduled observations. Stat Med. 2000; 19: 1303-1318.

39. Lin H, Turnbull BW, McCulloch CE, Slate EH. Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. Journal of the American Statistical Association. 2002; 97: 53-65.

40. Muthén B, Brown HC. Estimating drug effects in the presence of placebo response: causal inference using growth mixture modeling. Stat Med. 2009; 28: 3363-3385.

41. Bauer DJ, Curran PJ. Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. Psychol Methods. 2003; 8: 338-363.

42. Connell AM, Frye AA. Growth mixture modelling in developmental psychology: Overview and demonstration of heterogeneity in developmental trajectories of adolescent antisocial behaviour. Infant and Child Development. 2006; 15: 609-621.

43. Li F, Barrera M Jr, Hops H, Fisher KJ. The longitudinal influence of peers on the development of alcohol use in late adolescence: a growth mixture analysis. J Behav Med. 2002; 25: 293-315.

44. Ram N, Grimm KJ. Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. International Journal of Behavioral Development. 2009; 33: 565-576.

45. Hix-Small H, Duncan TE, Duncan SC, Okut H. A multivariate associative finite growth mixture modeling approach examining adolescent alcohol and marijuana use. Journal of Psychopathology and Behavioral Assessment. 2004; 26: 255-270.

46. Wang X, Schumitzky A, D'Argenio DZ. Nonlinear Random Effects Mixture Models: Maximum Likelihood Estimation via the EM Algorithm. Comput Stat Data Anal. 2007; 51: 6614-6623.

47. McLachlan GJ, Peel D. Robust cluster analysis via mixtures of multivariate t-distributions. Advances in pattern recognition: Springer. 1998; 658-666.

48. Peel D, McLachlan GJ. Robust mixture modelling using the t distribution. Statistics and computing. 2000; 10: 339-348.

49. Andrews JL, McNicholas PD, Subedi S. Model-based classification via mixtures of multivariate t-distributions. Computational Statistics & Data Analysis. 2011; 55: 520-529.

50. Shoham S. Robust clustering by deterministic agglomeration EM of mixtures of multivariate< i> t-distributions. Pattern Recognition. 2002; 35: 1127-1142.

51. Lin TI, Lee JC, Ni HF. Bayesian analysis of mixture modelling using the multivariate t distribution. Statistics and Computing. 2004; 14: 119-30.

52. Jiao S, Zhang S. The t-mixture model approach for detecting differentially expressed genes in microarrays. Funct Integr Genomics. 2008; 8: 181-186.

53. McNicholas PD, Subedi S. Clustering gene expression time course data using mixtures of multivariate< i> t-distributions. Journal of Statistical Planning and Inference. 2012; 142: 1114-1127.

54. Shoham S, Fellows MR, Normann RA. Robust, automatic spike sorting using mixtures of multivariate t-distributions. J Neurosci Methods. 2003; 127: 111-122.

55. Nguyen TM, Wu QM. Robust Student's-t mixture model with spatial constraints and its application in medical image segmentation. IEEE Trans Med Imaging. 2012; 31: 103-116.

56. Ho RT, Fong TC, Cheung IK. Cancer-related fatigue in breast cancer patients: factor mixture models with continuous non-normal distributions. Qual Life Res. 2014; 23: 2909-2916.

57. Sfikas G, Nikou C, Galatsanos N, editors. Robust image segmentation with mixtures of student's t-distributions. Image Processing, 2007 ICIP 2007 IEEE International Conference on. 2007.

58. Lin TI, Lee JC, Yen SY. Finite mixture modelling using the skew normal distribution. Statistica Sinica. 2007; 17: 909-927.

59. Lin TI. Maximum likelihood estimation for multivariate skew normal mixture models. Journal of Multivariate Analysis. 2009; 100: 257-265.

60. Cabral CRB, Lachos VH, Prates MO. Multivariate mixture modeling using skew-normal independent distributions. Computational Statistics & Data Analysis. 2012; 56: 126-142.

61. Basso RM, Lachos VH, Cabral CRB, Ghosh P. Robust mixture modeling based on scale mixtures of skew-normal distributions. Computational Statistics & Data Analysis. 2010; 54: 2926-2941.

62. Frühwirth-Schnatter S, Pyne S. Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. Biostatistics. 2010; 11: 317-336.

63. Arellano-Valle RB, Genton MG, Loschi RH. Shape mixtures of multivariate skew-normal distributions. Journal of Multivariate Analysis. 2009; 100: 91-101.

64. Lin TI, Lee JC, Hsieh WJ. Robust mixture modeling using the skew t distribution. Statistics and Computing. 2007; 17: 81-92.

65. Lin TI. Robust mixture modeling using multivariate skew t distributions. Statistics and Computing. 2010; 20: 343-356.

66. Pyne S, Hu X, Wang K, Rossin E, Lin TI, Maier LM, et al. Automated high-dimensional flow cytometric data analysis. Proc Natl Acad Sci USA. 2009; 106: 8519-8524.

67. Vrbik I, McNicholas P. Analytic calculations for the EM algorithm for multivariate skew-t mixture models. Statistics & Probability Letters. 2012; 82: 1169-1174.

68. Fernandes E, Pacheco A, Penha-Gonçalves C. Mapping of quantitative trait loci using the skew-normal distribution. J Zhejiang Univ Sci B. 2007; 8: 792-801.

69. Zou Y, Zhang Y. Use of skew-normal and skew-t distributions for mixture modeling of freeway speed data. Transportation Research Record: Journal of the Transportation Research Board. 2011; 2260: 67-75.

70. Pewsey A. Modelling asymmetrically distributed circular data using the wrapped skew-normal distribution. Environmental and Ecological Statistics. 2006; 13: 257-269.

71. Lee SX, McLachlan GJ. On mixtures of skew normal and skew t-distributions. Advances in Data Analysis and Classification. 2013; 7: 241-266.

72. Lee S, McLachlan GJ. Finite mixtures of multivariate skew t-distributions: some recent and new results. Statistics and Computing. 2014; 24: 181-202.

73. Lee SX, McLachlan GJ. Model-based clustering and classification with non-normal mixture distributions. Statistical Methods & Applications. 2013; 22: 427-454.

74. Lee S, McLachlan GJ. On the fitting of mixtures of multivariate skew t-distributions via the EM algorithm. arXiv preprint arXiv:11094706. 2011.

75. Wang K, Ng SK, McLachlan GJ, editors. Multivariate skew t mixture models: applications to fluorescence-activated cell sorting data. Digital Image Computing: Techniques and Applications, 2009 DICTA'09; 2009: 526-531.

76. Rossin E, Lin TI, Ho HJ, Mentzer SJ, Pyne S. A framework for analytical characterization of monoclonal antibodies based on reactivity profiles in different tissues. Bioinformatics. 2011; 27: 2746-2753.

77. Riggi S, Ingrassia S. Modeling high energy cosmic rays mass composition data via mixtures of multivariate skew-t distributions. arXiv preprint arXiv:13011178. 2013.

78. Karlis D, Santourian A. Model-based clustering with non-elliptically contoured distributions. Statistics and Computing. 2009; 19: 73-83.

79. Subedi S, McNicholas PD. Variational Bayes approximations for clustering via mixtures of normal inverse Gaussian distributions. Advances in Data Analysis and Classification. 2014; 8: 167-193.

80. Cabral CRB, Bolfarine H, Pereira JRG. Bayesian density estimation using skew student-t-normal mixtures. Computational Statistics & Data Analysis. 2008; 52: 5075-5090.

81. Franczak B, Browne R, McNicholas P. Mixtures of shifted asymmetric Laplace distributions. 2014; 36: 1149-1157.

82. Browne RP, McNicholas PD. A mixture of generalized hyperbolic distributions. arXiv preprint arXiv:13051036. 2013.

83. Muthén B, Asparouhov T. Growth mixture modeling: Analysis with non-Gaussian random effects. Longitudinal data analysis. 2008: 143-165.

84. Contreras-Reyes JE, Arellano-Valle RB. Growth curve based on scale mixtures of skew-normal distributions to model the age-length relationship of Cardinalfish (Epigonus Crassicaudus). arXiv preprint arXiv:12125180. 2012.

85. Muthén B, Asparouhov T. Growth mixture modeling with non-normal distributions. Stat Med. 2015; 34: 1041-1058.

86. Lu X, Huang Y. Bayesian analysis of nonlinear mixed-effects mixture models for longitudinal data with heterogeneity and skewness. Stat Med. 2014; 33: 2830-2849.

87. Lu X, Huang Y, Zhu Y. Finite mixture of nonlinear mixed-effects joint models in the presence of missing and mismeasured covariate, with application to AIDS studies. Computational Statistics & Data Analysis. 2014.

88. Huang Y, Yan C, Yin P, Lu M. A mixture of hierarchical joint models for longitudinal data with heterogeneity, non-normality, missingness and covariate measurement errors. Journal of Biopharmaceutical Statistics. 2015.

89. Huang Y, Cheng F, Qiu H. Simultaneous Bayesian inference on a finite mixture of mixed-effects Tobit joint models for longitudinal data with multiple features. Journal of Applied Statistics.

90. Huang Y, Chen J, Yin P. Hierarchical mixture models for longitudinal immunologic data with heterogeneity, non-normality, and missingness. Stat Methods Med Res. 2014.

91. Huang Y, Yan C. Piecewise mixture modeling for longitudinal virologic data with heterogeneity, non-normality and missingness. Statistics in Biopharmaceutical Research.

92. Huang Y, Qiu H, Yan C. Semiparametric mixture modeling for skewed-longitudinal data: A Bayesian approach. Annals of Biometrics & Biostatistics. 2015; 2: 1011.

93. Huang Y. A Bayesian mixture of semiparametric mixed-effects joint models for skewed-longitudinal and time-to-event data. Statistics in Medicine.

94. Lo Y, Mendell NR, Rubin DB. Testing the number of components in a normal mixture. Biometrika. 2001; 88: 767-778.

95. Akaike H. A new look at the statistical model identification. Automatic Control, IEEE Transactions on. 1974; 19: 716-723.

96. Schwarz G. Estimating the dimension of a model. The annals of statistics. 1978; 6: 461-464.

97. Sclove SL. Application of model-selection criteria to some problems in multivariate analysis. Psychometrika. 1987; 52: 333-343.

98. Jedidi K, Ramaswamy V, DeSarbo WS. A maximum likelihood method for latent class regression involving a censored dependent variable. Psychometrika. 1993; 58: 375-394.

99. McLachlan GJ, Peel D, Basford KE, Adams P. The EMMIX software for the fitting of mixtures of normal and t-components. Journal of Statistical Software. 1999; 4: 1-14.

100. Muthén LK, Muthén BO. Mplus. Statistical analyses with latent variables. User's guide. 1998.

101. Stutz J, Cheeseman P. Auto Class-a Bayesian approach to classification. In Maximum entropy and Bayesian methods. 1996; 70: 117-126.

102. Wolfe JH. NORMIX: Computational methods for estimating the parameters of multivariate normal mixtures of distributions (no. npra-srm-68-2). Naval personnel research activity san diego calif. 1967.

103. Macdonald P. MIX software for mixture distributions. 1988.

104. Fraley C, Raftery A, Fraley MC. The mclust package. 2007.

105. Benaglia T, Chauveau D, Hunter D, Young D. Mixtools: An r package for analyzing finite mixture models. Journal of Statistical Software. 2009; 32: 1-29.

106. Leisch F. FlexMix: A general framework for finite mixture models and latent glass regression in R. 2004; 11: 1-18.

107. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - a Bayesian modeling framework: concepts, structure, and extensibility. Statistics and Computing. 2000; 10: 325-337.